Proportional Size Distribution (PSD)

Alex J. Benecke

January 10, 2018

I will use a Chi-Square test to see if there is a difference between PSD between years (2013 - 2017).

Data Preparation

Load Data

```
lmbs <- read.csv("Data/Clean-Data/2012-2017_nearshore-survey-largemouth-bass_Stock_CLEAN.csv") %>%
    filter(Year >= 2013) %>% arrange(Year, FID, Length)
lmbs$fyr <- as.factor(lmbs$fyr)
```

```
unique(lmbs$Year) ### See that there is no 2012
```

```
## [1] 2013 2014 2015 2016 2017
```

View Data

```
(lmbs.LF <- xtabs(~Year+gcat,data=lmbs))
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      41    41
##   2014        18      57    65
##   2015        15      38    14
##   2016        11      44    52
##   2017         9      15    11
```

Chi-Squares Test

Is there a difference in the number of fish in each gabelhouse length category (PSD-X) during the years 2013 - 2017?

```
chisq.test(lmbs.LF)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmbs.LF
## X-squared = 20.055, df = 8, p-value = 0.01013
```

This seems to suggest that the proportional stock distribution (PSD) is different for largemouth bass between years ($X^2 = 20.055$, df = 8, P = 0.01013).

Where are the differences in PSD X-Y for each year?

Creating a table of percent of fish in each gabelhouse length interval (PSD X-Y) in each year.

```
round(prop.table(lmbs.LF,margin=1)*100,0)
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      42    42
##   2014        13      41    46
##   2015        22      57    21
##   2016        10      41    49
##   2017        26      43    31
```

Remarkably the percent of quality fish is the same for 2014 and 2016. the percent of quality length fish is very similar between 2013 (42%), 2014 (41%), 2016 (41%), and 2017 (43%). This trend is almost the same for stock length fish (2013 42%, 2014 46%, 2016 49%), however 2015 and 2017 have a smaller percentage of stock length fish (21% and 31% respectivley). This may be explained by more variability among years for this length category, reduced sampling efficiency for smaller fish, or unstable recruitment. The year 2015 seems to break the trend with a greater percentage of preferred (22%) and quality (57%) length individuals than other years.

1) Could this be some sort of sampling bias?
2) Could this be a result of sampling different sites? where the 2015 sites more suitable for LMB?

Compare PSD-Q between years 2013 - 2017

```
lmbs %<>% mutate(gcatQ=mapvalues(gcat,
                                 from=c("stock","quality","preferred"),
                                 to=c("quality-", "quality+", "quality+")),
                gcatQ=droplevels(gcatQ))

(lmb.LFQ <- xtabs(~Year+gcatQ,data = lmbs))
```

```
##       gcatQ
## Year   quality+ quality-
##   2013       57       41
##   2014       75       65
##   2015       53       14
##   2016       55       52
##   2017       24       11
```

```
chisq.test(lmb.LFQ)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmb.LFQ
## X-squared = 16.815, df = 4, p-value = 0.0021
```

```
(ps.Q <- c(chisq.test(lmb.LFQ[c(1,2),])$p.value,  ### 2013-2014
           chisq.test(lmb.LFQ[c(1,3),])$p.value,  ### 2013-2015
           chisq.test(lmb.LFQ[c(1,4),])$p.value,  ### 2013-2016
           chisq.test(lmb.LFQ[c(1,5),])$p.value,  ### 2013-2017
           chisq.test(lmb.LFQ[c(2,3),])$p.value,  ### 2014-2015
           chisq.test(lmb.LFQ[c(2,4),])$p.value,  ### 2014-2016
           chisq.test(lmb.LFQ[c(2,5),])$p.value,  ### 2014-2017
           chisq.test(lmb.LFQ[c(3,4),])$p.value,  ### 2015-2016
           chisq.test(lmb.LFQ[c(3,5),])$p.value,  ### 2015-2017
           chisq.test(lmb.LFQ[c(4,5),])$p.value)) ### 2016-2017
```

```r
(p.val.Q <- p.adjust(ps.Q))
```

```
##  [1] "13-14" "13-15" "13-16" "13-17" "14-15" "14-16" "14-17" "15-16"
##  [9] "15-17" "16-17"
```

```
##     Year p-value Adjusted p
## 1  13-14  0.5694     1.0000
## 2  13-15  0.0084     0.0675
## 3  13-16  0.4060     1.0000
## 4  13-17  0.3781     1.0000
## 5  14-15  0.0007     0.0064
## 6  14-16  0.8338     1.0000
## 7  14-17  0.1583     0.9500
## 8  15-16  0.0005     0.0046
## 9  15-17  0.3515     1.0000
## 10 16-17  0.1144     0.8007
```

The PSD-Q of largemouth bass is different for at least one of the years during 2013 - 2017 (Chi-Squared, $X^2 = 16.815$, df $= 4$, p $= 0.0021$). The adjusted p-values show a *significant difference* in PSD-Q between years 2014 - 2015 (p $= 0.0064$) and 2015 - 2016 (p $= 0.0046$). The PSD-Q is not different between any other years. However 2013 and 2015 may be different (p $= 0.0675$).

Compare PSD-P between years 2013 - 2017

```r
lmbs %<>% mutate(gcatP=mapvalues(gcat,
                                 from=c("stock","quality","preferred"),
                                 to=c("preferred-", "preferred-", "preferred+")),
                 gcatP=droplevels(gcatP))

(lmb.LFP <- xtabs(~Year+gcatP,data = lmbs))
```

```
##       gcatP
## Year   preferred+ preferred-
##   2013         16         82
##   2014         18        122
##   2015         15         52
##   2016         11         96
##   2017          9         26
```

```r
chisq.test(lmb.LFP)
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  lmb.LFP
## X-squared = 8.2649, df = 4, p-value = 0.08234
```

```r
(ps.P <- c(chisq.test(lmb.LFP[c(1,2),])$p.value,  ### 2013-2014
           chisq.test(lmb.LFP[c(1,3),])$p.value,  ### 2013-2015
           chisq.test(lmb.LFP[c(1,4),])$p.value,  ### 2013-2016
           chisq.test(lmb.LFP[c(1,5),])$p.value,  ### 2013-2017
           chisq.test(lmb.LFP[c(2,3),])$p.value,  ### 2014-2015
           chisq.test(lmb.LFP[c(2,4),])$p.value,  ### 2014-2016
           chisq.test(lmb.LFP[c(2,5),])$p.value,  ### 2014-2017
```

```r
          chisq.test(lmb.LFP[c(3,4),])$p.value,  ### 2015-2016
          chisq.test(lmb.LFP[c(3,5),])$p.value,  ### 2015-2017
          chisq.test(lmb.LFP[c(4,5),])$p.value)) ### 2016-2017
```

```
## Warning in chisq.test(lmb.LFP[c(4, 5), ]): Chi-squared approximation may be
## incorrect
```

```r
(p.val.P <- p.adjust(ps.P))
```

```
##  [1] "13-14" "13-15" "13-16" "13-17" "14-15" "14-16" "14-17" "15-16"
##  [9] "15-17" "16-17"
```

```
##     Year p-value Adjusted p
## 1  13-14  0.5724     1.0000
## 2  13-15  0.4378     1.0000
## 3  13-16  0.2837     1.0000
## 4  13-17  0.3329     1.0000
## 5  14-15  0.1212     0.8485
## 6  14-16  0.6716     1.0000
## 7  14-17  0.1048     0.8387
## 8  15-16  0.0498     0.4565
## 9  15-17  0.8964     1.0000
## 10 16-17  0.0456     0.4565
```

The PSD-P of largemouth bass is not different for any years during 2013 - 2017 (Chi-Squared, $X^2 = 8.26$, df = 4, p = 0.08). The adjusted p-values show no difference in the PSD-P between years (2013 - 2017). T

PSD-X Without 2017

I want to re-run the analysis without the year 2017 because I have decided to throw this year out due to incomplete sampling. I have left the code with 2017 above for future reference.

View Data

```r
(lmbs.LF <- xtabs(~Year+gcat,data=lmbs[lmbs$Year<2017,]))
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      41    41
##   2014        18      57    65
##   2015        15      38    14
##   2016        11      44    52
```

Chi-Squares Test (no 2017)

Is there a difference in the number of fish in each gabelhouse length category (PSD-X) during the years 2013 - 2017?

```r
chisq.test(lmbs.LF)
```

```
##
##  Pearson's Chi-squared test
##
```

```
## data:  lmbs.LF
## X-squared = 16.694, df = 6, p-value = 0.01048
```

This seems to suggest that the proportional stock distribution (PSD) is different for largemouth bass between years ($X^2 = 16.694$, df = 8, P = 0.01048).

Where are the differences in PSD X-Y for each year?

Creating a table of percent of fish in each gabelhouse length interval (PSD X-Y) in each year.

```
round(prop.table(lmbs.LF,margin=1)*100,0)
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      42    42
##   2014        13      41    46
##   2015        22      57    21
##   2016        10      41    49
```

Remarkably the percent of quality fish is the same for 2014 and 2016. the percent of quality length fish is very similar between 2013 (42%), 2014 (41%), and 2016 (41%). This trend is almost the same for stock length fish (2013 42%, 2014 46%, 2016 49%), however, 2015 has a smaller percentage of stock length fish (21%). This may be explained by more variability among years for this length category, reduced sampling efficiency for smaller fish, or unstable recruitment.

1) Could this be some sort of sampling bias?
2) Could this be a result of sampling different sites? where the 2015 sites more suitable for LMB?

Compare PSD-Q between years 2013 - 2016

```
lmbs.1316 <- lmbs[lmbs$Year<2017,]
lmbs.1316 %<>% mutate(gcatQ=mapvalues(gcat,
                                 from=c("stock","quality","preferred"),
                                 to=c("quality-", "quality+", "quality+")),
                 gcatQ=droplevels(gcatQ))

(lmb.LFQ.1316 <- xtabs(~Year+gcatQ,data = lmbs.1316))
```

```
##       gcatQ
## Year   quality+ quality-
##   2013       57       41
##   2014       75       65
##   2015       53       14
##   2016       55       52
```

```
chisq.test(lmb.LFQ.1316)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmb.LFQ.1316
## X-squared = 15.306, df = 3, p-value = 0.001573
```

```
(ps.Q.1316 <- c(chisq.test(lmb.LFQ.1316[c(1,2),])$p.value,  ### 2013-2014
           chisq.test(lmb.LFQ.1316[c(1,3),])$p.value,  ### 2013-2015
           chisq.test(lmb.LFQ.1316[c(1,4),])$p.value,  ### 2013-2016
           chisq.test(lmb.LFQ.1316[c(2,3),])$p.value,  ### 2014-2015
```

```r
           chisq.test(lmb.LFQ.1316[c(2,4),])$p.value,  ### 2014-2016
           chisq.test(lmb.LFQ.1316[c(3,4),])$p.value)) ### 2015-2016

(p.val.Q.1316 <- p.adjust(ps.Q.1316))
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
##    Year p-value Adjusted p
## 1 13-14  0.5694     1.0000
## 2 13-15  0.0084     0.0337
## 3 13-16  0.4060     1.0000
## 4 14-15  0.0007     0.0036
## 5 14-16  0.8338     1.0000
## 6 15-16  0.0005     0.0027
```

The PSD-Q of largemouth bass *is significantly different* for at least one of the years during 2013 - 2016 (Chi-Squared, $X^2 = 15.306$, df = 3, p = 0.0015). The adjusted p-values show a *significant difference* in PSD-Q between years 2013 - 2015 (p = 0.0337), 2014 - 2016 (p = 0.0036), and 2015 - 2016 (p = 0.0027). The PSD-Q *is not significantly different* between any other years.

Compare PSD-P between years 2013 - 2016

```r
lmbs.1316 %<>% mutate(gcatP=mapvalues(gcat,
                                  from=c("stock","quality","preferred"),
                                  to=c("preferred-", "preferred-", "preferred+")),
                 gcatP=droplevels(gcatP))

(lmb.LFP.1316 <- xtabs(~Year+gcatP,data = lmbs.1316))
```

```
##       gcatP
## Year   preferred+ preferred-
##   2013         16         82
##   2014         18        122
##   2015         15         52
##   2016         11         96
```

```r
chisq.test(lmb.LFP.1316)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmb.LFP.1316
## X-squared = 5.4469, df = 3, p-value = 0.1418
```

```r
(ps.P.1316 <- c(chisq.test(lmb.LFP.1316[c(1,2),])$p.value,  ### 2013-2014
           chisq.test(lmb.LFP.1316[c(1,3),])$p.value,  ### 2013-2015
           chisq.test(lmb.LFP.1316[c(1,4),])$p.value,  ### 2013-2016
           chisq.test(lmb.LFP.1316[c(2,3),])$p.value,  ### 2014-2015
           chisq.test(lmb.LFP.1316[c(2,4),])$p.value,  ### 2014-2016
           chisq.test(lmb.LFP.1316[c(3,4),])$p.value))  ### 2015-2016

(p.val.P.1316 <- p.adjust(ps.P.1316))
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
##    Year p-value Adjusted p
```

```
## 1 13-14  0.5724       1.000
## 2 13-15  0.4378       1.000
## 3 13-16  0.2837       1.000
## 4 14-15  0.1212       0.606
## 5 14-16  0.6716       1.000
## 6 15-16  0.0498       0.299
```

The PSD-P of largemouth bass is *not significantly different* for any years during 2013 - 2016 (Chi-Squared, $X^2 = 5.45$, df = 3, p = 0.14). The adjusted p-values show no difference in the PSD-P between years (2013 - 2016).